

Enhancing Data Quality with Paradata

George C. Jones III
U.S. Census Bureau

Overview

- Focal Points of Managing Data Collection
- American Community Survey (ACS)
- High Workloads and Data Quality
 - Reducing Interviewer Effects
 - High Workload Report
- Item Nonresponse and Data Quality
 - Item Nonresponse Report

Managing Data Collection

- Response Rates
- Cost/Budget
- Data quality
 - Most challenging to manage
 - The biggest obstacle to overcome, not only in managing data quality but managing field interviewers, is the lack of direct information about what takes place during the interview.

American Community Survey (ACS)

- Nationwide survey that collects and produces information on demographic, social, economic, and housing characteristics
- 50 states and Puerto Rico
- Housing Units (HU) and Group Quarters (GQ)
- Data Collection:
 - 3 month data collection cycle
 - PAPI, CAWI, CATI, and CAPI

High Workload Report

- Quality control review of assignments that identifies interviewers with large workloads
- Issues with interviewers and high workloads
 - Poor data quality
 - Increase in the risk of data falsification
 - Morale ----> Turnover
 - Vacancies
 - Interviewer effects

Reducing Interviewer Effects

- Any influence on data collected from interviewing people that is caused by the attributes, behavior, or characteristics of the individual conducting the interview
- Can have substantial impact on survey data
- More interviewers conducting fewer interviews reduces interviewer effects
- Dilutes the influence of individual interviewers

High Workload Report

Interviewer	Survey Period	Day of Survey	Current Workload	Cases Checked In	Response Rate	Worst Case Response Rate
Jane Doe	201504 ACS HU CAPI	20	41	40	85.00%	80.95%
Bob Voe	201504 ACS HU CAPI	20	40	40	95.16%	95.16%
Bill Poe	201504 ACS HU CAPI	20	46	46	93.34%	93.34%
Will Coe	201504 ACS HU CAPI	20	47	47	92.02%	92.02%
Tom Roe	201504 ACS HU CAPI	20	41	35	93.33%	66.67%
Rita Voe	201504 ACS HU CAPI	20	46	46	94.36%	94.36%
Jan Roe	201504 ACS HU CAPI	20	48	48	76.47%	76.47%
Karen Poe	201504 ACS HU CAPI	20	47	47	90.48%	90.48%
Mary Coe	201504 ACS HU CAPI	20	44	32	95.45%	61.76%

Item Nonresponse

- Identifies selected key variables, or questions, for which the interviewer recorded that the respondent's answer was “don't know” or “refused”
- Missing values compromise data quality
- Two versions of this report
 - By Geography
 - By Interviewer

Item Nonresponse Report

Survey Period	SSF	FS	Interviewer	# Required Responses	% Nonresponses	% DK	% RF	# Nonresponses	# DK	# RF	Current Workload	Cases Checked In
201505 ACS HU CAPI	94	9482	John Doe	25	76.0%	0.0%	76.0%	19	0	19	13	9
	91	9188	Jane Doe	205	51.7%	41.0%	10.7%	106	84	22	18	11
	98	9886	Bob Voe	34	50.0%	2.9%	47.1%	17	1	16	3	3
	91	9188	Bill Poe	215	46.5%	44.2%	2.3%	100	95	5	28	15
	94	9484	Will Coe	33	45.5%	6.1%	39.4%	15	2	13	22	13
	95	9587	Tom Roe	40	45.0%	40.0%	5.0%	18	16	2	10	3
	98	9891	Rita Voe	325	42.8%	8.9%	33.8%	139	29	110	23	10
	92	9286	Jan Roe	76	42.1%	2.6%	39.5%	32	2	30	9	4
	95	9489	Karen Poe	48	41.7%	4.2%	37.5%	20	2	18	14	7
	98	9891	Mary Coe	64	40.6%	7.8%	32.8%	26	5	21	15	8

Challenges

- Managing interviewers remotely from a Regional Office
- Getting interviewers to transmit daily
- Finding the balance between rewarding good interviewers with more work and the maximum amount of work an interviewer can handle without compromising data quality (High Workload Report)

Thank you.

george.c.jones@census.gov